1. Introduction

In fictitious domain methods (see [5] or for more recent work [2,7,8]) one is often faced with the choice of either integrating the equations over the whole computational mesh, i.e. also in the non-physical part, or only integrate within the physical domain. In the first case the method is robust, but inaccurate due to the lack of consistency. Methods using the second approach, on the other hand, are accurate, but the condition number of the finite element matrix depends on how the domain boundary cuts the mesh. If the cut results in elements with very small intersections with the physical domain, the system matrix may be very ill-conditioned, as we show below.

In this Note we will propose a simple trick that allows to enhance robustness of the method without sacrificing accuracy. The idea is to add a penalty term in the interface zone that extends the coercivity of the physical domain to all of the elements intersected by the domain boundary, also the part where the solution has no physical significance. Herein we only discuss the application of the method in the framework of fictitious domain methods, but it can also be used to enhance robustness in extended finite element methods (see [1] for an application in elasticity), unfitted methods and Chimera methods.

We restrict the discussion to Poisson’s problem:

\[-\Delta u = f \quad \text{in } \Omega,\]
\[u = 0 \quad \text{on } \partial \Omega,\]

(1)
where $\Omega$ is some open connected subset of $\mathbb{R}^2$ with smooth or polygonal boundary $\partial \Omega$. Below $c$ and $C$ will denote generic constants that may change at each occurrence, but that always are independent of the local mesh size $h$ and the boundary/mesh intersection.

2. Finite element framework

Let $\{T_h\}$ be a family of shape regular and (for simplicity) quasi uniform, triangulations without hanging nodes, such that $T_h = \{T\}$, with mesh function $h(x)$ such that $h(x)|_T = \text{diam}(T)$. For all $T_h$ we assume (i) $\bar{\Omega} \subset \Omega_T := \bigcup_{T \in T_h} T$; (ii) $\Omega_T \setminus \partial \Omega \neq \emptyset$; (iii) $T \cap \Omega \neq \emptyset \forall T \in T_h$. Further assume that $\partial \Omega$ and $T_h$ satisfy assumptions [A2]–[A3] of [6] and no element contains more than one corner of $\partial \Omega$, essentially implying that the boundary is resolved by the mesh. Define the space of continuous piecewise polynomial functions on the mesh $T_h$ by $V^k_h := \{v_h \in H^1(\Omega_T) : v_h|_{T} \in P_k(T) \forall T \in T_h\}$, $N_V := \dim V^k_h$. To illustrate the theory we consider the following non-symmetric fictitious domain method inspired by Nitsche’s method [9] for the approximation of (1): find $u_h \in V^k_h$ such that

$$a_h(u_h, v_h) = \int_{\Omega} f v_h \, dx \quad \forall v_h \in V^k_h,$$

where $a_h(u_h, v_h) := (\nabla u_h, \nabla v_h)_\Omega - (\nabla u_h \cdot n, v_h)_{\partial \Omega} + (\nabla v_h \cdot n, u_h)_{\partial \Omega} + (\gamma h^{-1} u_h, v_h)_{\partial \Omega}, \gamma \geq 0$ with $n$ the outward pointing unit normal to $\partial \Omega$. The forms $(\cdot, \cdot)_{\partial \Omega}$ and $(\cdot, \cdot)_\Omega$ denote the $L^2$-scalar products over $\Omega$ and $\partial \Omega$ respectively, with induced norms $\|u\|_{0,\partial \Omega} = (u, u)_{\partial \Omega}^{\frac{1}{2}}$ and $\|u\|_{0,\Omega} = (u, u)_{\Omega}^{\frac{1}{2}}$. We also introduce the norms

$$\|u_h\|_{0,\Omega} := \|\nabla u_h\|_{0,\Omega}^2 + \|h^{-\frac{1}{2}} u_h\|_{0,\partial \Omega}^2 \quad \text{and} \quad \|u_h\|_{0,\partial \Omega} := \|\nabla u_h\|_{0,\partial \Omega}^2 + \|h^{-\frac{1}{2}} u_h\|_{0,\partial \Omega}^2.$$ 

The following Poincaré-type inequalities hold for all $u \in H^1(\Omega)$ and $u \in H^1(\Omega_T)$ respectively

$$\|u\|_{0,\Omega} \leq C_P \|u\|_{1,\Omega} \quad \text{and} \quad \|u\|_{0,\partial \Omega} \leq C_P \|u\|_{1,\partial \Omega}.$$ 

On $V^k_h$ we also have the inverse inequalities

$$\|u_h\|_{1,\partial \Omega} \leq C h_{\text{min}}^{-\frac{1}{2}} \|u_h\|_{0,\partial \Omega}, \quad \|u_h\|_{0,\partial \Omega} \leq C h_{\text{min}}^{-\frac{1}{2}} \|u_h\|_{0,\partial \Omega} \quad \text{where} \quad h_{\text{min}} = \min_{T \in T_h} h_T.$$ 

The formulation (2) then satisfies the coercivity and continuity relations (using Lemma 3 of [6])

$$a_h(u_h, v_h) \geq a_h(u_h, v_h), \quad |a_h(u_h, v_h)| \leq M_a \|u_h\|_{1,\partial \Omega} \|v_h\|_{1,\partial \Omega}.$$ 

Both $a_h$ and $M_a$ are independent of the interface position. Note however that $a_h(u_h, v_h)$ may not be shown to be continuous on $\|\cdot\|_{1,\partial \Omega}$ independently of the interface position.

Let $A$ denote the system matrix associated to (2) defined by $A := \{A_{ij}\}_{i,j=1}^{N_V}$ and $A_{ij} := a_h(\phi_i, \phi_j)$ where $\{\phi_i\}_{i=1}^{N_V}$ denotes the nodal basis functions of $V^k_h$. We define the condition number of $A$ by $\kappa(A) := \|A\| \|A^{-1}\|$, $\|A\| := \sup_{x \in \mathbb{R}^{N_V}} |A|_{x|_{\mathbb{R}^{N_V}}}$, where $|\cdot|_{\mathbb{R}^{N_V}}$ denotes the Euclidean norm on $\mathbb{R}^{N_V}$. Further let $M$ denote the mass matrix defined by the bilinear form $(u_h, v_h)_{\Omega_T}$. Since $T_h$ is a standard conforming mesh on the domain $\Omega_T$ we have the estimates

$$\mu_{\min} \frac{1}{2} \|u\|_{N_V} \leq \|u_h\|_{0,\Omega_T} \leq \mu_{\max} \|u\|_{N_V},$$

where $u := \{u_i\}_{i=1}^{N_V} \in \mathbb{R}^{N_V}$ and $u_h$ is the associated function in $V^k_h$ defined by $u_h := \sum_{i=1}^{N_V} u_i \phi_i$. Here $\mu_{\min}$ and $\mu_{\max}$ denote the smallest and largest eigenvalues of $M$ (see [4]).

3. Reason for poor condition number of the system matrix

Under assumption (iii) the system matrix must be regular. It will however become ill-conditioned if for some node $x_0$ the set of triangles containing $x_0$ has small intersection with $\Omega$ as we now show.

**Lemma 3.1.** Consider a fixed mesh $T_h$. Assume that for some node $x_0$ meas$_d(\bigcup_{T, x_0 \in T} T \cap \Omega) < \epsilon$ and meas$_{d-1}(\bigcup_{T, x_0 \in T} T \cap \partial \Omega) < \epsilon$. Then there exists a constant $c_0 > 0$ that may depend on the mesh parameter $h$, but not on $\epsilon$, such that $\kappa(A) \geq \frac{c_0}{\epsilon^2}$.

**Proof.** First pick $x \in \mathbb{R}^{N_V}$, with associated $x_0 \in V^k_h$, such that supp($x_0$) $\subset \bar{\Omega}$ and therefore $\|x_0\|_{0,\Omega} = \|x_0\|_{0,\partial \Omega}$. Then by the coercivity of $a_h(\cdot, \cdot)$, the Poincaré inequality (3) and the relation (7), we have

$$\|A\| \geq \frac{|Ax|_{N_V}}{|x|_{N_V}} \geq \frac{(Ax, x)_{N_V}}{|x|_{N_V}^2} \geq \frac{a_h(x_0, x_0)}{|x|_{N_V}^2} \geq \frac{\alpha}{\epsilon} \|x_0\|_{0,\partial \Omega}^2 \geq \alpha C_P \mu_{\min}.$$
Consider now \( \|A^{-1}\| \). By definition \( \|A^{-1}\| = \sup_{v \in \mathbb{R}^N} \frac{|A^{-1}v|_N}{|v|_N} = \sup_{u \in \mathbb{R}^N} \frac{|u|_N}{|Au|_N} \). However, also by definition \( |Au|_N = \sup_{w \in \mathbb{R}^N} \frac{|Aw|_N}{|w|_N} \) and \( |Au|_N = \sup_{w \in \mathbb{R}^N} \frac{a(u, w)}{|w|_N} \). Now pick \( u_h = \phi_h \), then \( |u|_N = 1 \). Since by assumption \( \operatorname{meas}_d(\operatorname{supp}(\phi_h) \cap \Omega) < \epsilon \) and \( \operatorname{meas}_{d-1}(\operatorname{supp}(\phi_h) \cap \partial \Omega) < \epsilon \) we deduce that \( |(\nabla \phi_h, \nabla w_h)_{\Omega}| \leq C_\epsilon |\nabla|_2^{-1} \|w_h\|_{\Omega} \) and \( |(\nabla \phi_h \cdot n, w_h)_{\partial \Omega} + (\nabla w_h \cdot n, \phi_h)_{\Omega}| \leq C_\epsilon |\nabla|_2^{-1} \|w_h\|_{\Omega} \). Hence we have the upper bound \( |a(u_h, w_h)| \leq C_\epsilon h^{-2} |w_h|_{\Omega} \), leading to \( |Au|_N \leq C_\epsilon h^{-2} \|w_h\|_{\Omega} \), from which the claim follows, since then \( \|A^{-1}\| \geq C_\epsilon^{-1} h^2 \epsilon^{-2} \). \( \square \)

### 4. The Ghost penalty method

The idea is to decompose the boundary zone of the mesh in \( N_P \) patches \( P_i \) with diameters \( h_{P_i} \approx O(h) \), consisting of a moderate number of elements, such that the condition number of finite element matrices is eliminated. The condition number of a boundary-fitted FEM method defined on a mesh is eliminated. The condition number of \( Ah \) reduces to a reference patch. But then it follows that

\[
\frac{\|\nabla \phi_h \cdot n, w_h \|_{\partial \Omega} + \|\nabla w_h \cdot n, \phi_h \|_{\Omega}}{|\nabla|_2^{-1} \|w_h\|_{\Omega} \leq C_\epsilon |\nabla|_2^{-1} \|w_h\|_{\Omega}}.
\]

We now show that this extended coercivity is sufficient to guarantee a uniform upper bound on the condition number

\[
\kappa(G) = \|G\| \equiv \sup_{v \in \mathbb{R}^N} \frac{|Gv|_N}{|v|_N}.
\]

Definition 4.1 (Ghost penalty). \( s_h(u_h, v_h) := \sum_{i=1}^{N_P} s_i(u_h, v_h) \), where \( s_i(u_h, v_h) := \int_{P_i} \frac{1}{h_{P_i}} (u_h - \pi_h u_h)v_h \, dx \).

It is easy to verify that the penalty term is symmetric, \( s_h(u_h, v_h) = s_h(v_h, u_h) \), using the orthogonality of the \( L^2 \)-projection. It is also clear from standard approximation results that the following continuity holds

\[
\|s_h(u_h, v_h) \|_{\Omega} \leq C_\epsilon |\nabla|_2^{-1} \|w_h\|_{\Omega} \] for all \( u_h, v_h \in L^2(\Omega) \).

Lemma 4.2. There exists \( \alpha_\epsilon > 0 \) such that, for all \( v_h \in V_h^k \),

\[
\|A_s\|_{\Omega} = \int_{\Omega} f v_h \, dx. \quad \forall v_h \in V_h^k.
\]

Proof. It is enough to prove that for all \( l \), there holds for some \( \alpha_\epsilon > 0 \), \( \alpha_\epsilon \|\nabla v_h\|_{\Omega}^2 \leq \|\nabla v_h\|_{\Omega}^2 \leq \|\nabla v_h\|_{\Omega}^2 + s_i(v_h, v_h) \). To this end map \( P_l \) to a reference patch \( \hat{P} \) with diameter of order 1 and the mapped function \( \hat{v} \). We let \( s_l(\hat{v}, \cdot) \) denote the mapped penalty operator. The result now follows from norm equivalence on discrete spaces. Observe that if \( s_l(\hat{v}, \cdot) = 0 \) then \( \hat{v} \in P_l \) and hence is a global polynomial on the patch. But then it follows that \( \|\hat{v}\|_{\hat{P}}^2 \leq \hat{C}_l \|\hat{v}\|_{\hat{P}}^2 + s_l(\hat{v}, \cdot) \), since a polynomial that vanishes on a set of non-zero measure is zero everywhere. The constant \( \hat{C}_l \) depends on \( \alpha_\epsilon \) and the polynomial order \( k \), but not on the interface position. The claim follows by scaling back to the physical patch. \( \square \)

We now show that this extended coercivity is sufficient to guarantee a uniform upper bound on the condition number of the matrix \( G := A + S \) defined by the bilinear form \( A_h(u_h, v_h) := a_h(u_h, v_h) + s_h(u_h, v_h) \). Note that this form satisfies a stronger coercivity than that of \( A_h(u_h, v_h) \) alone, since by Lemma 4.2 and (5) there exists \( \alpha_\epsilon > 0 \) such that \( \alpha_\epsilon \|v_h\|^2_{\Omega} \leq A_h(u_h, v_h) \leq C_A \|A_h\|_{\Omega} \|v_h\|. \)

Lemma 4.3. The condition number of the system matrix resulting from the formulation (8) satisfies the upper bound

\[
\kappa(G) \leq C_A h^{-2},
\]

where the constant \( C_A := M_A C_2 h^{-2} \|A\|_{\Omega} \) is independent of the boundary/mesh intersection \( \partial \Omega \cap T_h \).

Proof. By definition \( |Gv|_N = \sup_{w \in \mathbb{R}^N} \frac{|Gw|_N}{|w|_N} = A_h(u_h, w_h) \). Using now the continuity of \( A_h(\cdot, \cdot) \), an inverse inequality and the bound (7) we may write \( A_h(u_h, w_h) \leq M_A C_h \|u_h\|_{\Omega} h^{-1} \|w_h\|_{\Omega} \) and hence \( |Gv|_N \leq M_A C^2 h^{-2} \|v_h\|_{\Omega} \). Applying once again the inverse inequality (4) and the bound (7) we have \( |Gv|_N \leq M_A C^2 h^{-2} \|v_h\| \leq M_A C^2 h^{-2} \|v_h\| \), resulting in

\[
\kappa(G) \leq C_A h^{-2}.
\]
\begin{equation}
\|G\| = \sup_{u \in \mathbb{R}^{N_V}} \frac{|G|_{N_V}}{|u|_{N_V}} \leq M_A C_f^2 \mu_{\text{max}} h_{\text{min}}^{-2}.
\end{equation}

Similarly for the norm of the inverse we have, using coercivity and the Poincaré inequality
\begin{equation}
C_f^{-1} \alpha_A \mu_{\text{min}}^{-\frac{1}{2}} |u|_{N_V} \leq C_f^{-1} \alpha_A \|u_h\|_{0, \Omega_T} \leq \alpha_A \|u_h\|_{1, h, \Omega_T} \leq \frac{A_h(u_h, u_h)}{\|u_h\|_{1, h, \Omega_T}} = \frac{(G, u)_{N_V}}{\|u_h\|_{1, h, \Omega_T}} \leq \frac{|G|_{N_V} \mu_{\text{min}} C_f \|u_h\|_{1, h, \Omega_T}}{\|u_h\|_{1, h, \Omega_T}} = \mu_{\text{min}}^{-\frac{1}{2}} C_f |G|_{N_V}.
\end{equation}

Since \( u \) is arbitrary and \( G \) is regular it follows, by setting \( v = G u \), that
\begin{equation}
|G^{-1}| = \sup_{v \in \mathbb{R}^{N_V}} \frac{|G^{-1} v|_{N_V}}{|v|_{N_V}} \leq \sup_{G u \in \mathbb{R}^{N_V}} \frac{|u|_{N_V}}{|G u|_{N_V}} \leq \mu_{\text{min}}^{-1} C_f^2 \alpha_A^{-1}.
\end{equation}

We conclude by inserting the bounds (9) and (10) into the definition of \( \kappa(G) \). \( \square \)

**Remark 1.** If we let \( \mathcal{F}_{\Omega_T} \) denote the set of all interior faces in the set \( \bigcup_{i=1}^{N_P} \mathcal{P}_i \), then the distributed penalty term can be shown, using a scaling argument, to be upper bounded by a multi penalty term acting on the jumps of the normal derivatives of all orders over element boundaries: \( s_h(u_h, u_h) \leq c j(u_h, u_h) \), where \( j(u_h, v_h) = \sum_{F \in \mathcal{F}_{\Omega_T}} \sum_{k=1}^{k_F} \int_{F} h_F^{-\frac{1}{2}} |\delta^k_h u_h||\delta^k_h v_h| \, ds \).

Here \( \delta^k_h \) denotes the normal derivative of order \( k \), \( |x| \) denotes the jump of quantity \( x \) over the element face \( F \) and \( h_F \) the diameter of the face \( F \). This jump penalty term is another valid ghost penalty operator.

In order to prove optimal convergence of the numerical scheme (8) it is convenient to consider an extension of the exact solution, \( u \in H^{k+1}(\Omega) \) denoted \( E u \in H^{k+1}(\Omega_T) \). \( E u |_{\Omega} = u \) such that \( \|E u\|_{H^{k+1}(\Omega_T)} \leq c \|u\|_{H^{k+1}(\Omega)} \) (cf. [3]). If \( i_h : H^2(\Omega_T) \to V_h^1 \) denotes the standard Lagrange interpolation operator, it is straightforward to prove that \( s_h(i_h E u, i_h E u)^{1/2} \leq c(i_h E u, i_h E u)^{1/2} \leq \varepsilon_h \|u\|_{H^{k+1}(\Omega)} \). This weak consistency estimate is sufficient to prove optimal convergence of the method. As a final result we prove the optimal convergence in energy norm of the formulation (2).

**Proposition 4.4.** Assume that the problem (1) admits a solution \( u \in H^{k+1}(\Omega) \) and that \( u_h \) is the solution of (8). Then there holds \( \|u - u_h\|_{1, h, \Omega} \leq \varepsilon_h \|u\|_{H^{k+1}(\Omega)} \).

**Proof.** Let \( e_h := i_h E u - u_h \). Then by coercivity and Galerkin orthogonality
\begin{equation}
\alpha_A \|e_h\|_{1, h, \Omega_T}^2 \leq A_h(e_h, e_h) = a_h(i_h E u - u, e_h) + s_h(i_h E u, e_h) \leq (M_a \|i_h E u - u\|_{1, h, \Omega_T}^2 + \|h^{\frac{1}{2}} (i_h E u - u) \cdot n\|_{0, \partial \Omega_T}^2 + M_f^2 s_h(i_h E u, i_h E u)^{1/2}) \|e_h\|_{1, h, \Omega_T}^2.
\end{equation}

We conclude by noting that by Lemma 3 of [6], standard interpolation estimates and by the properties of the extension operator there holds \( \|i_h E u - u\|_{1, h, \Omega_T} + \|h^{\frac{1}{2}} (i_h E u - u) \cdot n\|_{0, \partial \Omega_T} \leq \varepsilon_h^k \|u\|_{H^{k+1}(\Omega)} \) and \( s_h(i_h E u, i_h E u)^{1/2} \leq \varepsilon_h \|u\|_{H^{k+1}(\Omega)} ^{1/2} \). \( \square \)

**References**


